The capsule and genetic background, rather than specific individual loci, strongly influence *in vitro* pneumococcal growth kinetics

Chrispin Chaguza[1,2,3,4‡], Daan W. Arends[5], Stephanie W. Lo[3], Indri Hapsari Putri[5], Anna York[1], John A. Lees[6], Anne L. Wyllie[1], Daniel M. Weinberger[1], Stephen D. Bentley[3], Marien I. de Jonge[5#], Amelieke J.H. Cremers[5,7‡#]

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, Connecticut, USA

[2]Yale Institute for Global Health, Yale University, New Haven, Connecticut, USA

[3]Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

[4]Department of Host-Microbe Interactions, St Jude Children's Research Hospital, Memphis, Tennessee, USA

[5]Laboratory of Medical Immunology, Radboud Institute of Molecular Life Sciences, Radboudumc Center for Infectious Diseases, Radboudumc, Nijmegen, The Netherlands

[6]European Molecular Biology Laboratory–European Bioinformatics Institute, Cambridge, United Kingdom

[7]Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

#Senior author

*These authors contributed equally

‡Correspondence: Chrispin.Chaguza@yale.edu (C.C.) and Amelieke.Cremers@radboudumc.nl (A.J.H.C.)

Abstract

Bacterial growth characteristics intrinsic to each strain can impact and influence gene expression, antibiotic susceptibility, and disease pathogenesis. However, little is known about specific genomic variations that influence these bacterial growth features. Here, we investigate the impact of *Streptococcus pneumoniae* genetics on its *in vitr*o growth features to shed light on genes that may be important targets in the development of vaccines and therapeutics. We use statistical models to estimate growth features and demonstrate that they varied significantly across capsular serotypes and lineages, were strongly correlated with phylogeny, and showed high heritability, highlighting a strong genetic basis. Despite this, genome-wide association studies revealed no specific genomic loci statistically associated with the growth features independently of the genetic background, including those in the locus responsible for capsular polysaccharide synthesis. Our findings suggest that the serotype and lineage, as well as a combination of genomic loci, influence intrinsic pneumococcal growth kinetics, which may have implications for pneumococcal disease pathogenesis.

Keywords

Streptococcus pneumoniae; *in vitro* growth; GWAS; population genomics; capsular serotype; genetic background

Introduction

Once termed "the captain of the men of death" by Sir William Osler in the 20th century, *Streptococcus pneumoniae*, or the pneumococcus, remains a significant cause of life-threatening invasive diseases, such as sepsis, pneumonia, and meningitis[1], despite the widespread implementation of life-saving serotype-specific polysaccharide pneumococcal conjugate vaccines (PCVs)[2]. The high capsular diversity, comprising ~100 serotypes with varying potential for colonising the nasopharynx and causing invasive diseases[3,4] and mortality[5–7], demonstrates the remarkable genetic diversity of the pneumococcus. Such high genetic diversity reflects the plasticity of the pneumococcal genome, which readily exchanges genetic material with other strains through recombination and horizontal gene transfer[8]. Capsule switching, resulting from the swapping of the capsule biosynthesis (cps) loci, which encode the extracellular polysaccharide capsule that determines the capsular serotype, exemplifies genetic exchange between pneumococcal strains[9]. While serotypes typically exist in specific lineages[10], such as those defined by the pneumococcal global sequence cluster (GPSC) nomenclature[11], some serotypes are expressed across distinct lineages, while some lineages express multiple serotypes, reflecting capsule switching[12]. However, the impact of individual capsular serotypes, genetic background, and their combinations on the bacterial phenotypes, especially those associated with disease pathogenesis and clinical manifestations, remains less understood.

The pneumococcus possesses an arsenal of virulence factors. The extracellular polysaccharide capsule is a primary virulence factor crucial for pneumococcal transmissibility, survival, and pathogenicity in the human host[13–15]. Recent studies on the pneumococcus and other bacterial pathogens suggest that variability in growth kinetics affects disease pathogenesis, clinical manifestation, and antibiotic therapy. For example, differential pneumococcal growth features appear to correlate with fitness for colonisation and invasive disease,[16] clinical manifestation[16–18] and response to environmental conditions[18–20]. Studies of other bacterial pathogens suggest that growth rate influences gene and protein expression[21,22], adhesion[23], bacterial competition[24], plasmid replication[25], bacteriophage dynamics[26], and antibiotic accumulation and efficacy[27–29]. Similarly, experimental and mathematical modelling studies have demonstrated that bacterial growth correlates with bacterial load and host-mediated cell lysis, impacting dispersal to

different host cells and tissues during systemic infection[30]. Previous studies of pneumococcus and other bacteria have suggested the influence of environmental factors[19,20] and bacterial-specific factors, including capsular serotype[18,31] and the presence of plasmids[32] on bacterial growth phenotypes. These growth phenotypes may consequently impact the pathogenesis of invasive pneumococcal disease[16–18]. However, the specific pneumococcal genomic variations affecting growth kinetics remain less understood.

Studying bacterial growth kinetics sheds light on the influence of genes, genomic variation, and growth media conditions on bacterial fitness[33,34]. However, studying bacterial growth kinetics *in vivo* remains intrinsically challenging due to the complexity of the conditions in the hosts, although innovative approaches are being developed[35,36]. On the other hand, *in vitro* methods provide a valuable alternative for assessing the impact of specific bacterial and environmental factors on bacterial fitness. Although these *in vitro* methods have yielded remarkable insights into bacterial fitness, they have primarily provided strain-specific rather than population-level insights due to the analysis of a small number of isolates from a limited number of genetic backgrounds.

Here, we studied the impact of pneumococcal genetics on *in vitro* growth kinetics to understand the influence of bacterial genetic variation on pneumococcal pathogenesis. Specifically, we performed high-throughput quantification of the *in vitro* growth kinetics and whole genome sequencing of 348 invasive pneumococcal isolates from the Netherlands to determine the impact of genomic variation on growth kinetics. We hypothesised that the capsular serotype, genetic background, and specific individual genomic loci independent of the genetic background influence in vitro pneumococcal growth kinetics features or parameters derived from the growth curves. We tested this hypothesis through a series of statistical and comparative genomics analyses, including a bacterial genome-wide association study (GWAS) approach. Our study provides insights into the influence of capsular serotype, genetic background or lineage, and individual genomic loci on pneumococcal growth kinetics – a fundamental trait that potentially influences nasopharyngeal colonisation, invasive disease pathogenesis, and clinical manifestations.

Results

Population structure of a diverse collection of Dutch invasive pneumococcal isolates and derivation of their growth kinetics

We aimed to understand the impact of the capsular serotype, genetic background, and specific genomic loci, independent of the clonal population structure, on the pneumococcal growth features **(Fig. 1 and Supplementary Data 1)**. Therefore, we investigated the *in vitro* growth kinetics of 348 pneumococcal isolates collected from patients with invasive pneumococcal diseases via the Pneumococcal Bacteraemia Collection Nijmegen (PBCN) cohort study at two hospitals in Nijmegen, the Netherlands before (2000-2006) and after (2007-2011) the introduction of the seven-valent pneumococcal conjugate vaccine (PCV) [PCV7] into the Dutch paediatric immunisation programme[37]. Our isolates represent 34 capsular serotypes, 40 pneumococcal lineages, as defined by the global pneumococcal sequence cluster (GPSC) nomenclature[11], and 99 sequence types (STs) specified by the pneumococcal multilocus sequence typing (MLST) scheme[38]. An interactive and annotated maximum likelihood phylogeny of the isolates used in this study is available online at the Microreact platform for open data visualisation and sharing for genomic epidemiology (https://microreact.org/project/wxmdblfgprepwugvbbtaq7). We estimated a Simpson diversity index of 0.932 (95% confidence interval [CI]: 0.924–0.940), 0.937 (95% CI: 0.928–0.945), and 0.958 (95% CI: 0.949–0.966) for the serotypes, lineages, and STs, respectively. The high richness and diversity of pneumococcal serotypes and lineages highlighted a substantial population-level diversity of the study isolates in the Netherlands, consistent with earlier studies of the same population[39,40].

Capsular serotype and genetic background significantly impact pneumococcal growth kinetics

To explore the impact of capsular serotype and genetic background or lineage on the pneumococcal growth features estimated from a fitted growth curve, namely lag-phase duration, maximum growth density (H_{max}), maximum growth change (ΔH), and average growth rate (r) **(Fig. 1, 2, and Supplementary Data 1),** we assessed the relationship between these growth

features and serotypes and lineages associated with the study isolates **(Fig. 3a)**. In general, we found significant variability of the growth features across serotypes: average growth rate, maximum growth density, maximum growth change, and lag phase duration ($P<0.0001$; Kruskal-Wallis test) consistent with previous studies[18,19] **(Fig. 3b-e)**. We observed similar patterns across different pneumococcal lineages **(Fig. 3f-i and Supplementary Data 2)**.

Recombination and horizontal gene transfer may lead to the replacement of part or the entire capsule biosynthesis locus, resulting in the production of a polysaccharide capsule distinct from the original one, a phenomenon called capsule-switching[41]. Consequently, some lineages are associated with multiple serotypes, and similarly, a single serotype can be expressed in distinct genetic backgrounds[12]. Examples of lineages containing multiple serotypes included GPSC3, GPSC4, and GPSC7, and likewise, serotypes, including 3, 19A, and 19F, were expressed in multiple genetic backgrounds, which also express other serotypes **(Fig. 3a)**. Overall, the average growth rate for serotypes expressed in different genetic backgrounds broadly showed similar patterns. For example, lineages expressing serotype 14 had consistently higher average growth rates. In contrast, lineages expressing serotype 3 showed lower rates, consistent with the notion that capsule production interferes with pneumococcal growth[18] **(Fig. 3b).** Therefore, we investigated the impact of the capsular serotype and genetic background on the estimated pneumococcal growth features.

First, we identified lineages that expressed multiple serotypes and serotypes found across multiple lineages. To ensure a robust statistical comparison, we only focused on serotype-lineage combinations with at least three sequenced isolates in our study. We identified two lineages, GPSC3 and GPSC7, with sufficient sequenced isolates for comparison, which expressed multiple serotypes **(Fig. 3j)**. We observed higher growth rates for serotype 33F than serotype 8 ($P=0.0007$; Kruskal-Wallis test).

In contrast, we found no statistically significant differences in growth rates between serotypes 23A and 23F in lineage GPSC7 ($P=0.2568$; Kruskal-Wallis test). We speculated that the absence of statistically significant differences between serotypes 23A and 23F reflected their presence in the same serogroup; therefore, they had similar capsule properties. Interestingly, we observed

similar patterns for the maximum growth density, maximum growth change, and lag phase duration features between serotypes on the GPSC3 and GPSC7 lineages **(Fig. 3k-m and Supplementary Data 2)**. Next, we investigated the growth rates of the same serotypes expressed in distinct genetic backgrounds. Although some serotypes exhibited similar growth rates across distinct genetic backgrounds, such as serotype 3 (P=0.4599; Kruskal-Wallis test), others showed different growth rates across different lineages, as seen with serotypes 14 (P=0.0006; Kruskal-Wallis test) and 19F (P=0.0105; Kruskal-Wallis test). We noted similar patterns for the other growth features: maximum growth density, maximum growth change, and lag phase duration **(Fig. 3n-q)**.

These findings demonstrate that the capsular serotype and genetic background significantly impact pneumococcal growth kinetics. The resultant growth features likely depend on the interaction between these factors, whereby some combinations may result in more pronounced growth effects than others.

Pneumococcal growth kinetics show heritability and correlation with the phylogeny of the isolates

Based on the findings that both capsular serotype and genetic background are associated with growth kinetics, we hypothesised that the estimated pneumococcal growth features exhibit a high correlation between the pneumococcal growth features and the phylogeny, i.e., phylogenetic signal. To assess this, we first performed ancestral state reconstruction to map the derived pneumococcal growth features or parameters across the whole genome phylogeny of the isolates. By reconstructing the estimated growth features on the internal nodes of the phylogeny, we found high but variable associations between the growth features and different genetic backgrounds **(Fig. 4, Supplementary Fig. 2 and 3, and Supplementary Data 2)**. Broadly, genetically similar isolates from the same phylogenetic lineages or genetic backgrounds exhibited similar growth kinetics, varying in degree depending on the specific growth feature. Altogether, these findings suggest a strong phylogenetic signal in the growth features.

Next, we calculated the phylogenetic signal of each pneumococcal growth feature using Pagel's λ statistic38, a metric typically ranging from 0 to 1 that evaluates the correlation between given traits and the phylogeny, with higher values indicating a stronger phylogenetic signal. The amount of the phylogenetic signals based on the normal transformed growth features was highest for the average growth rate (Pagel's λ=0.72, 95% CI: 0.61 to 0.81; $P<2.08\times10^{-64}$), maximum growth change (Pagel's λ=0.60, 95% CI: 0.40 to 0.76; $P<1.53\times10^{-16}$), and maximum growth density (Pagel's λ=0.73, 95% CI: 0.57 to 0.84; $P<4.18\times10^{-26}$), while the lag phase duration showed a lower phylogenetic signal (Pagel's λ=0.30, 95% CI: 0.17 to 0.47; $P<5.42\times10^{-16}$) **(Fig. 5a)**. These findings suggest that there is a heterogeneous correlation of the growth features with the phylogeny.

To quantify the variability in pneumococcal growth kinetics explained by genotype, we calculated the narrow-sense heritability (h^2) for the four normal-transformed estimated growth features and the maximum growth rate from Arends et al.[17]. We found strong, though not equal, evidence for a genetic basis for all the estimated growth features: average growth rate (h^2=0.76, 95% CI: 0.68 to 0.85), maximum growth change (h^2=0.80, 95% CI: 0.71 to 0.89), maximum growth density (h^2=0.88, 95% CI: 0.83 to 0.94), and lag phase duration (h^2=0.29, 95% CI: 0.14 to 0.43) **(Fig. 5b, Supplementary Data 2)**. To further understand the impact of the capsular serotype and genetic background, we fitted a linear mixed effects model for each growth parameter with serotype and lineage (defined by the GPSC nomenclature[11]) as random effects to estimate the amount of variance of the growth features explained by the serotype (h^2_{serotype}) and lineage or GPSC (h^2_{GPSC}) calculated as the ratio of the variance of the serotype and lineage components relative to the total variance, including the residual or environmental effects. We calculated h^2_{serotype}=0.47 and h^2_{GPSC}=0.11 for the average growth rate, h^2_{serotype}=0.13 and h^2_{GPSC}=0.28 for the maximum growth change, h^2_{serotype}=0.29 and h^2_{GPSC}=0.46 for the maximum growth, h^2_{serotype}=0.08 and h^2_{GPSC}=0.39 for the lag phase duration feature.

Altogether, these findings demonstrate a substantial impact of serotype and lineage on the variation in pneumococcal growth features, whereby most serotypes tend to have similar characteristics regardless of the genetic background. Similarly, some genetic backgrounds tend

to exhibit similar phenotypic patterns regardless of the serotypes, but the serotype appears to have a more significant influence on the growth characteristics.

No individual genomic loci are associated with pneumococcal growth features

After finding moderate to high narrow-sense heritability of the growth features, we then undertook a GWAS analysis to identify specific genomic loci associated with these features, independent of the strains' genetic background using a linear mixed-effects model. We performed separate GWAS analyses of the maximum growth change, maximum growth density, maximum growth change, and lag phase duration phenotypes derived from the growth curve. We conducted the GWAS using variable-length *k*-mer sequences or unitigs to capture genetic variation in the pneumococcal genomes at different levels, including SNPs, multi-allelic sites, and insertions and deletions[43,44]. We found no individual unitigs associated with the derived growth feature phenotypes **(Fig. 6a)**. Furthermore, repeating the GWAS analyses adjusting for the serotype revealed no statistically significant hits associated with the derived growth phenotypes, suggesting that the variants in the pneumococcal strains are correlated with the growth kinetics independent of the serotype **(Fig. 6b)**. Overall, the observed and expected statistical significance values in the QQ-plots for our study revealed no apparent issues with the population structure of the isolates, which is a significant confounder in similar studies[45] **(Fig. 6c, d)**. While the capsule is biologically relevant for pneumococcal growth (Fig. 3b-q), the absence of individual loci even within the capsule biosynthesis locus with elevated signals associated with the growth features may be a mere consequence of adjusting for the population structure as pneumococcal serotypes are typically associated with specific lineages (Fig. 3a). Overall, these findings suggest that the contribution of individual genomic loci in influencing pneumococcal growth kinetics does not stand out, potentially reflecting the polygenic nature of the phenotype and cumulative effect of the genetic background or lineage.

Average *in vitro* growth rates are not strongly correlated with a serotype's invasiveness or its prevalence in carriage or disease

Previous studies have elucidated that bacterial growth rate may influence colonisation and disease[20], including systemic dissemination during infection[30] and disease manifestations[17,31].

Pneumococcal strains that are prolific growers may exhibit better survival and replication in systemic tissues than those that are slow growers. We used isolates collected before the introduction of the PCV7 vaccine in the Netherlands[46] to assess the association between *in vitro* growth rate of the isolates and their population-level characteristics: prevalence in invasive disease and nasopharyngeal carriage and invasiveness, i.e., the odds of being isolated in invasive disease compared to the carriage state. Our first hypothesis was that the average *in vitro* growth rate for each pneumococcal serotype is positively associated with the prevalence of that serotype in invasive diseases. Our analysis revealed no evidence of an association between the prevalence of serotypes in invasive disease and the *in vitro* growth rate (Spearman correlation: -0.36, P=0.098) **(Fig. 7a, Supplementary Data 3)**. Second, we hypothesized that the average *in vitro* growth rate of the serotypes is positively associated with the prevalence of the serotypes in the nasopharyngeal carriage, as fitter strains would grow faster than less fit strains. Similarly, we found no association between the prevalence of serotypes in carriage and the *in vitro* growth rate (Spearman correlation=-0.41, P=0.12) **(Fig. 7b, Supplementary Data 3)**. We repeated the analysis after excluding serotype 3 and found no evidence of an association between growth rate and carriage prevalence (Spearman correlation=-0.28, P=0.314), which suggests that serotype 3 alone was driving the previously observed association. Finally, we tested the hypothesis that the growth rate is positively associated with the invasiveness of the serotypes. We found no association between the invasiveness and the *in vitro* growth rates of the serotypes (Spearman correlation: 0.144, P=0.594) **(Fig. 7c and Supplementary Data 3)**. These findings suggest that growth rates alone may not be sufficient to explain the prevalence and invasiveness of pneumococcal serotypes, indicating the influence of other pneumococcal factors, microbiome (e.g., viral infection), and host-related factors in determining pneumococcal disease susceptibility.

Discussion

Bacterial growth kinetics impact several traits, including gene and protein expression[21,22], antibiotic accumulation and efficacy[27–29], cell surface adhesion[23], intra- and inter-specific competition[24], response to environmental conditions[18–20], plasmid replication[25], bacteriophage dynamics[26], colonisation and disease pathogenesis[16], including clinical manifestation[16–18]. Whilst previous studies have attempted to understand bacterial, host, and environmental factors

influencing bacterial growth, a major drawback has been a narrow focus on a small subset of strains[47]. Therefore, the insights gained from these studies have been mostly limited to specific strains, as they ignored potential variability across different genetic backgrounds. Such population-level studies are critical for genetically diverse bacterial pathogens with highly structured populations[11], substantial antigenic diversity[48], and recombination rates, such as the pneumococcus[8,10,41]. Here, we investigate the association between pneumococcal genetics and growth kinetics using a combination of high-throughput quantification of *in vitro* pneumococcal growth, whole-genome sequencing, and population genomic analyses employing a bacterial GWAS approach. Notably, our findings reveal a strong influence of the capsular serotype compared to the genetic background, and the absence of specific individual genomic loci modulating pneumococcal growth kinetics, particularly the *in vitro* growth rate. These findings may partially explain why some serotypes and lineages colonise and cause invasive diseases more effectively than others and may have implications for developing measures to prevent and control pneumococcal infections.

Our findings suggest that the capsular serotype and genetic background influence the growth kinetics of pneumococci. The findings are generally consistent with results from experimental studies by others[17–19], which used capsule-switched strains to exclude the impact of the genetic background and demonstrated an alteration in growth dynamics due to the capsule. However, there may be potential differences in the growth media and conditions. Specifically, encapsulated strains exhibited higher growth rates than unencapsulated strains in nutrient-rich conditions[18]. These findings suggest that the capsule is crucial for pneumococcal growth, but its effect depends on nutrient availability – promoting growth when nutrients are abundant and interfering with growth when nutrients are limited, respectively[18,31]. A recent study by Tóthpál et al.[19] suggested that the serotype partially explains pneumococcal growth and suggested the genetic backgrounds, which vary in their ability to colonise the nasopharynx and cause invasive diseases[4,49–51], may also play a role. However, the authors did not fully quantify the impact of the genetic background as the isolates were not sequenced[19]. Our study confirms the strong influence of both the pneumococcal capsular serotype and genetic background or lineage on the growth kinetics of the isolates. The influence of the serotype appeared to be greater than that of the genetic background on the average growth rate, while the other growth features seemed to show

an opposite trend. These findings suggest that the respective influence of the capsular serotype and lineage varies by specific derived pneumococcal growth features.

Interestingly, although our findings revealed that the capsular serotype and genetic background influence pneumococcal growth kinetics, statistical analysis showed that the phenotypic variability attributed to the capsular serotype was greater than that attributed to the genetic background. Specifically, the capsular serotype exhibits a more pronounced growth rate, maximum growth change, and maximum growth density than the genetic background. In contrast, the genetic background had a greater effect on the lag phase duration than the serotype, a growth phase suggested to be responsible for bacterial adaptation to new environments[52]. We acknowledge that our comparison of growth features in the same genetic background may also indicate a significant amount of genetic variation for some of the lineages, some of which could be responsible for changes in growth kinetics. These findings illustrate the combined, albeit variable, importance of the capsular serotype and genetic background on pneumococcal growth kinetics, which may partially explain the pathogenesis and prevalence of the pneumococcus in nasopharyngeal carriage and invasive disease.

Previous studies have revealed individual bacterial genomic loci that impact other phenotypes, including pathogenicity[53–61], virulence[62–65], host, niche, and environmental adaptation[56,58,66], conjugation[67], and antimicrobial resistance[68–75], independent of the genetic background. Although our study focused primarily on the pneumococcal growth rate parameter, the other growth parameters are also biologically relevant. For example, the lag phase duration potentially reflects the ability to adapt to environmental conditions, such as nutrient availability or stress, and may be more influenced by regulatory genes involved in sensing and responding to environmental cues[52,76]. On the other hand, the maximum growth rate may be impacted by genes controlling the rate of cell division under optimal conditions[77] while the maximum growth density, which reflects the final cell density, might be shaped by factors related to nutrient utilization efficiency, waste tolerance, or quorum sensing. The duration of the stationary phase may reflect the initiation of the lytic growth phase, possibly due to the effects of an autolysin inhibitor[78]. Our GWAS analysis, undertaken to gain a more nuanced understanding of how genetic variants contribute to different physiological aspects of microbial growth, revealed no

elevated signals in specific genomic loci associated with pneumococcal growth features. This suggests that individual variants may not significantly influence pneumococcal growth kinetics independently of the genetic background and serotype. Previous studies have shown that mutations in other capsule biosynthesis genes (*wchA* or *cpsE*) encoding a protein that links activated glucose phosphates to lipid carriers[48] also influence the growth dynamics of the pneumococcus[47]. However, our results contradict our initial speculation that genetic variation in the capsule polysaccharide synthesis genes would be strongly associated with pneumococcal growth features. These findings suggest that the contribution of specific individual pneumococcal genomic loci to growth kinetics does not stand out, hinting at the potential polygenic nature of growth kinetics and an overall impact of the genetic background, which exerts a cumulative effect, diluting the influence of individual loci. Uncovering such effects will require follow-up studies with larger sample sizes.

Compared to the findings of a previous study by Hathaway et al.[18], we found no robust evidence for an association between the *in vitro* growth rates and the population-level prevalence of the serotypes in nasopharyngeal carriage. However, this may have occurred as a result of a limitation of our study, since we studied a collection of randomly sampled invasive isolates, and the growth experiments were conducted under conditions similar to those of blood. In contrast, Hathaway et al.[18] used a non-random selection of both carriage and disease isolates. Isolates likely have adaptations to the environmental conditions in which they were isolated; therefore, *in vitro* growth characteristics for any given isolate may be different under experimental conditions simulating different niches. In addition, the *in vitro* growth environment cannot fully represent *in vivo* conditions and fails to consider the impact of selective pressures arising from serotype-specific responses (antibodies) or non-specific responses (phagocytosis) as well as potential interactions with other pathogens. The lack of association between the invasiveness and prevalence of serotypes in invasive disease and carriage, as well as the growth rates, could reflect that the conditions in the *in vitro* growth environment do not fully represent *in vivo* conditions. Collectively, these findings suggest that these and other factors may substantially influence the complex pathogenesis of pneumococcal disease *in vivo*.

The strength of this study lies in the analysis of an extensive collection of well-phenotyped natural isolates, combined with phenotyping and the estimation of growth features using statistical models and statistical genetics to determine the association between the growth features and genetic variation across the entire genome. As previously noted by Tothpal et al.[19], the culturing history of the isolates may impact the growth kinetics. In our study, we consistently cultured all isolates after isolation from patients using standardised culturing procedures (single assay, identical initial concentration, environmental conditions (temperature), and equipment to measure optical density, and repeated measurements to spot technical variability[17]). Compared to the present study, future studies would benefit from larger sample sizes and may yield more results, as we only explained a portion of the heritability of the growth features. We also acknowledge that the methods used to estimate heritability may also sometimes perform poorly on certain bacterial species; therefore, more robust methods are needed to validate our findings further[79]. We conducted our in vitro culture experiments for 15 hours, as the growth curves indicate that most isolates had reached the stationary phase by that time. Increasing the culturing time from 15 to >24 hours, as done elsewhere14,15, might be considered for follow-up studies. However Streptococcus pneumoniae undergoes autolysis upon reaching a specific cell density, which could distort growth measurements and complicate interpretation if incubation were prolonged. Furthermore, since the pneumococcus invades different niches and systemic tissues under varying conditions, including oxygen concentration and temperature[16,18–20], future studies should investigate the effect of these environmental conditions on pneumococcal growth and identify pneumococcal genomic variants associated with growth kinetics for each condition. Considering that our dataset contained only a third of the >100 known pneumococcal serotypes globally[3]; due to the geographical variation in the distribution of pneumococcal serotypes[80], future studies should investigate all serotypes to provide further insights into the impact of different serotypes on pneumococcal growth kinetics.

Additionally, considering that we generated the phylogeny of the entire pneumococcal species rather than isolates belonging to a single strain or lineage, as defined by several approaches[11,81,82], we did not remove recombinations using widely used methods designed for clonal bacterial populations, such as Gubbins[83]. Therefore, there is a possibility that some branch lengths in the phylogeny used for ancestral trait reconstruction analysis may be less accurate,

although the topology is usually robust to recombination[84]. Furthermore, there may be other variants that potentially influence the pneumococcal growth kinetics, which are regarded as not statistically significant in this study based on our conservative genome-wide significance threshold. As more similar studies are conducted, particularly those utilising much larger datasets, potential variants associated with pneumococcal growth kinetics may be discovered.

Our analysis of population-level genomic and high-throughput *in vitro* pneumococcal growth data suggests that the capsular serotype and lineage influence pneumococcal growth. However, although some growth features are highly heritable, there appears to be no specific genetic variants that strongly influence these phenotypes, suggesting that a combination of genomic loci with minor effects, individually or collectively, may play a greater role in pneumococcal growth kinetics. Overall, our study provides a proof-of-concept for combining phenotyping and population genomics to understand the genetic basis for pneumococcal growth, an approach which can be applied to study other bacteria of clinical significance and those used in bioengineering, genetics, and the food industry[85–87]. Ultimately, we anticipate that applying our approach will have implications for devising strategies to reduce life-threatening bacterial infections.

Methods

Ethics statement

The isolates used in this work were collected through an observational cohort study Pneumococcal Bacteraemia Collection Nijmegen (PBCN)[37]. The non-applicability of the Medical Research Involving Human Subjects Act (WMO) on the study design was confirmed by the Regional Ethics Committee, and the local medical ethics committees of the participating hospitals in the Netherlands approved the study procedures. Study procedures were approved by the Medical Ethical committees of the participating hospitals, including a waiver for individual informed consent (file number: 2020–6644 Radboudumc).

Sample characteristics and whole-genome sequencing

Three hundred forty-eight (n=348) pneumococcal isolates collected from patients with invasive pneumococcal disease through the PBCN observational cohort study[37] with in vitro growth data were available for the study. We included a single isolate per patient as invasive pneumococcal infections are typically clonal, so a single isolate is sufficient. The consecutive blood isolates were collected at two hospitals in Nijmegen, the Netherlands, before (2000-2006) and after (2007-2011) the introduction of the PCV7 vaccine into the Dutch paediatric immunisation programme[37]. Most of the bacteremia cases (97%) concerned adults. Additional information on the PBCN study, including collected clinical variables and phenotypic in vitro growth data, has been previously described[17,37,39,40]. To generate whole genome sequences, we sequenced a single isolate per patient using an Illumina HiSeq sequencer (Illumina, San Diego, CA, USA). We then assembled the generated reads using SPAdes genome assembler (version 3.14.0)[88], and evaluated the draft assemblies using assembly-stats (version: 1.0.1) (https://github.com/sanger-pathogens/assembly-stats). In addition, we used a combination of metrics, including genome size and the number of contigs, to assess the quality of the draft assemblies for inclusion in the analysis.

Measurement of *in vitro* pneumococcal growth kinetics and derivation of growth features

We measured the pneumococcal growth kinetics for slightly over 15 hours as described by Arends et al.[17]. In brief, we incubated a frozen aliquot of standardised pneumococcal inoculum in 1.5 ml of blood-like medium (50% M17, 50% casamino acids tryptone [CAT] medium, 0.25% glucose). We measured growth kinetics at 37°C and 5% CO_2 every 10 minutes for over 15 hours at an OD_{620} using a humidity cassette with a microplate reader (Spark 10M, Tecan, Switzerland). We added 15 µL of inoculum was added to 1.5 mL of prewarmed rich growth medium (45% M17, 45% CAT, 0.225% glucose, 10% fetal calf serum [Greiner Bio-one], 26 U/mL catalase [Sigma-Aldrich C1345]) in each well of a sterile flat-bottomed 48-well plate (Nunclon Surface, Nunc, Denmark). The growth medium was supplemented with catalase. We selected these nutrient-rich conditions to highlight the differences in the growth features of the pneumococcus. We performed six repeat measurements on three separate days for each isolate. To understand the growth kinetics of the invasive Dutch pneumococcal isolates, we fitted a modified logistic growth curve function, i.e. $\square_0 + \square_{\square\square\square}/(1 + \square^{-\square(\square-\square_0)})$, separately to the

growth data for each isolate and replicate, using the "optim" function in stats (version 4.3.2) R package (https://www.R-project.org/). We then averaged the inferred values for each replicate to obtain average estimates for each parameter. The parameter $L_{0\ \text{represents}}$ the initial cell density measured as absorbance or optical density at a wavelength of 620 nm (OD_{620}); L_{max} is the maximum cell density; t_0 is the initial time, and r is the average growth rate. We fitted two curves in total for each isolate based on the averaged data from the six replicates using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) optimisation algorithm by minimising the residual sum of squares[89]. We fitted the first curve to all the data points and the second curve to the data measured until the end of the stationary phase or the beginning of the death phase, whichever came earlier. We defined the end of the stationary phase or the start of the death phase as the time when the optical cell density decreased by more than 5% from the maximum cell density. The curve fitted based on the data measurements collected up to the end of the stationary phase or death phase, or the end of the experiment, if none of these phases had been reached, captured the growth dynamics better than the curve fitted using all the data points to accurately estimate the growth rate and the maximum cell density corresponding to the peak of the pneumococcal growth curve, especially for growth curves with observably short stationary phases and rapid autolysis. Therefore, parameters for the downstream analyses were inferred from the curves fitted to the reduced dataset.

We determined four growth phenotypes for each pneumococcal isolate estimated from the fitted growth curve, namely the lag-phase duration, maximum growth density (H_{max}), i.e., captured by parameter L_{max}, maximum growth change (ΔH) corresponding to the difference of the maximum and minimum density (L_{max}-L_0), and average growth rate (r). We defined the lag-phase duration as the time when the optical cell density was at least 1.5-fold higher than the initial cell density at the start of the experiment, as described by Arends et al.[17]. Due to the challenges of characterising the kinetics at the stationary or death phase, partly due to the high strain-to-strain variability or short study period, we were unable to derive the growth features that capture these phases, including the duration of the stationary and death phases. The logistic growth curves showed an excellent fit for the data, providing an opportunity to investigate the impact of capsular serotype and genetic background (or lineage), as well as specific genomic loci, independent of the genetic background, on the estimated growth parameters and features.

Molecular typing of sequenced isolates

Molecular typing of the isolates to determine capsular serotypes and sequence types (STs) was done *in silico* using Pathogenwatch (https://pathogen.watch/), a global platform for genomic surveillance[90]. We determined the capsular serotypes and STs using SeroBA (version v1.0.1)[91] and the multilocus sequence typing (MLST) scheme for the pneumococcus[38], respectively. Pneumococcal lineages were defined based on the global pneumococcal sequence cluster (GPSC) nomenclature[11] using PopPUNK (version 1.1.0)[92]. Additionally, we checked the species of each sequenced isolate using Speciator (version 3.0.1) in Pathogenwatch.

Construction of multiple sequence alignments and phylogenetic trees

We mapped the generated sequence reads for each isolate and a draft assembly of Streptococcus oralis strain SF100 genome (GenBank accession: NZ_CP069427) to the pneumococcal ATCC 700669 reference genome (GenBank accession: NC_011900) using Snippy (version 4.6.0) (https://github.com/tseemann/snippy). We identified 155,305 SNPs in the generated alignment using SNP-sites (version 2.3.2)[93] and then used them to construct a maximum-likelihood phylogeny using IQ-TREE (version 2.0.3)[94]. We used the alignment of variable sites by mapping to a reference genome to generate a maximum-likelihood phylogeny using IQ-TREE, as it has been shown to produce more accurate phylogenies[95]. The general time reversible (GTR) nucleotide substitution model and Gamma heterogeneity of the substitution rates among sites. We assessed branch support using the ultrafast bootstrap approximation specified by the "-B 1000" option[96]. We optimised bootstrapped trees using nearest neighbour interchange method on bootstrap alignments by selecting the "--bnni" option. We rooted the phylogenetic tree of the pneumococcal isolates using *Streptococcus oralis* strain SF100 genome as an outgroup (not shown in the phylogeny). An annotated and interactive version of the phylogeny is available on the Microreact web tool[97] (https://microreact.org/project/wxmdblfgprepwugvbbtaq7).

Quantifying the phylogenetic signal

To assess the correlation between the estimated growth features and pneumococcal phylogeny, we used models of continuous character evolution to estimate the phylogenetic signal using Pagel's λ[42] implemented in the "phylosig" function in phytools package (version 2.1.1)[98]. Pagel's λ transforms the internal branches relative to the terminal branches and generates values typically

ranging from zero to one, with smaller values indicating no phylogenetic signal or independent evolution of the phenotype from the phylogeny. In comparison, higher values close to one show a high phylogenetic signal or a decreased phenotype variability among genetically similar taxa at the tips of the phylogeny. First, we transformed the estimated growth features to a normal distribution using a rank-based inverse normal transform implemented in the RNOmni package (version 1.0.1.2) (https://cran.r-project.org/web/packages/RNOmni/). Next, we mapped the estimated continuous growth feature values onto the phylogeny using fast estimation of maximum likelihood for ancestral states[98].

Bacterial GWAS

We identified the presence and absence patterns of unitig sequences to capture genetic variation in the pneumococcal genomes included in this study for the GWAS analysis. We identified unitig sequences from 31bp *k*-mers using Bifrost (version 1.0.6)[99]. First, we inferred unitig sequences from a compacted De Bruijn graph constructed using *k*-mers of length 31, based on draft assemblies of all the pneumococcal isolates included in this study. We generated De Bruijn graphs for each isolate separately and then queried them with the unitig sequences found in the entire dataset to identify the presence and absence of each unitig found in each sequenced genome. We considered a unitig present when 100% of the *k*-mer was presented in the De Bruijn graph. We combined the presence and absence patterns of the unitig sequences from all the isolates to construct a single matrix.

Since the linear mixed effects models require the tested continuous phenotype to be normally distributed, we used the normalised growth features after applying the rank-based inverse normal transformation using the RNOmni package (version 1.0.1.2) (https://cran.r-project.org/web/packages/RNOmni/). We performed GWAS analysis of pneumococcal growth features using robust linear mixed-effects models implemented in Pyseer (version 1.3.7-dev)[100]. A kinship matrix specifying the variances and covariances of the random effects for each isolate was generated using the “similarity_pyseer” script in Pyseer. Using this kinship matrix, Pyseer fits an unobserved random effect for each isolate to account for the clonal population structure in the GWAS. We excluded unitigs with <5% minimum allele frequency and >95% maximum allele frequency. The inferred statistical significance for each variant was adjusted to correct for

multiple testing using the Bonferroni correction. We estimated the total number of unique unitig presence and absence patterns in the dataset using a script included in Pyseer (https://github.com/mgalardini/pyseer/blob/master/scripts/count_patterns.py). We used the obtained value when adjusting for multiple testing using the Bonferroni correction approach. Therefore, we defined the genome-wide statistical significance threshold as $P < 2.58\times10^{-07}$, i.e., □/G, where □=0.05 and G=194,053 is the number of unique unitig presence and absence patterns, while the suggestive threshold is $P<5.15\times10^{-06}$, whereby □=1. To check potential issues arising from the population structure, we generated quantile-quantile (QQ) plots from the output of GWAS analysis for each growth feature to compare the expected and observed *P*-values using qqman (version 0.1.7)[101]. The overall proportion of phenotypic variability explained by variation in the genome (narrow-sense heritability [h^2]) was estimated using GEMMA (version 0.98.1)[102] and Pyseer. The input files for GEMMA were prepared based on the unitig presence and absence data using PLINK (version 1.90b4) [103]. To identify variants associated with the growth features independently of the serotype, we repeated the GWAS by including the serotype as a covariate.

Annotation of variants identified by the GWAS analysis

To annotate the unitig sequences with statistical significance values below the suggestive and genome-wide significance thresholds, we used a custom BioPython [104] wrapper script for BLASTN (version 2.5.0+)[105] to identify genomic features in the ATCC 700669 reference genome (GenBank accession: NC_011900) and other complete pneumococcal genomes downloaded from GenBank[105]. The GWAS results were summarised and visualised using Manhattan plots in R (version 4.0.3) [https://www.R-project.org/].

Other statistical analysis

We compared the growth features between serotypes and lineages using the Kruskal-Wallis non-parametric test. The distribution of the growth features was plotted using violin plots using the vioplot package [version 0.4.0] (https://CRAN.R-project.org/package=vioplot). To assess the contribution of serotype and lineage to the variability in the pneumococcal growth features, we used linear mixed-effects models implemented in the lme4 package (version 1.1.35.1)[106] were used, with each growth feature specified as the outcome and the serotype and lineage as random effects. The statistical significance for the linear mixed model was inferred using the lmerTest

package [version 3.1.3] (https://CRAN.R-project.org/package=lmerTest). To assess the contribution of serotype and lineage to the variability of each growth feature, we calculated the proportion of variance attributable to the serotype and lineage components of the total variance. We calculated the average invasiveness of the serotypes based on a meta-analysis of data collected in different countries in infants before the introduction of the PCVs[50] using the random-effects model implemented in the metafor package [version 2.4.0] (https://CRAN.R-project.org/package=metafor). We defined invasiveness as the odds ratio of detecting a pneumococcal serotype among isolates causing invasive disease compared to nasopharyngeal carriage. We added a value of one to cells with zero counts. Due to the unavailability of the nasopharyngeal carriage data, we imputed the prevalence of serotypes in the carriage by dividing the prevalence in invasive disease by the average invasiveness of each serotype. Next, we used a Spearman correlation test to assess the association between the mean prevalence of serotypes in the nasopharyngeal carriage and invasive disease, and average invasiveness with the average *in vitro* pneumococcal growth rate.

Data availability

We have deposited the whole-genome sequencing data for the study isolates in the European Nucleotide Archive (ENA) and provided the accession numbers and other isolate metadata in Supplementary Data 1. All the other data supporting the findings of this study are described in this paper or are available as part of the supplementary material. We have also provided additional raw data and code used for the analysis on GitHub: https://github.com/ChrispinChaguza/SpnGrowthKinetics.

Acknowledgements

The authors would like to thank the study participants, guardians, and the clinical and laboratory staff who collected and processed the samples at various laboratories in the Netherlands. We would also like to acknowledge the support of the sequencing, core, parasites and microbes programme, the Bentley lab at the Wellcome Sanger Institute, and the Weinberger and Pitzer labs at the Yale School of Public Health for their feedback on the analysis. We acknowledge the participating hospitals and affiliated researchers for their contributions and support.

Funding information

This study was supported by the TARGET project (grant number: JPIAMR2019-087) funded by JPI-AMR–ZonMW, the Netherlands, and by the SNSF Swiss Postdoctoral Fellowship (SNSF grant number: TMPFP3_209768). Activities at the Wellcome Sanger Institute were funded by the Bill and Melinda Gates Foundation (grant number: OPP1034556) and Wellcome Trust (2016-2021 core award grant number: 206194). The funders had no role in study design, data collection, analysis, decision to publish, or manuscript preparation. The findings do not necessarily reflect the official views and policies of the author's institutions and funders. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

C.C., M.I.dJ., and A.J.H.C. conceived and designed the study. M.I.dJ., D.W.A., I.H.P., and A.J.H.C. collected and processed the pneumococcal isolates. S.W.L. and S.D.B. performed whole-genome sequencing, quality control, and genomic analysis. C.C. performed statistical and bioinformatic analyses. D.W.A. contributed to data analysis and interpretation. M.I.dJ. and A.J.H.C. contributed to data interpretation. C.C., D.W.A., M.I.dJ., and A.J.H.C. wrote the first draft of the manuscript. C.C., D.W.A., S.W.L., I.H.P., A.Y., J.A.L., A.L.W., D.M.W., S.D.B., M.I.dJ., and A.J.H.C. reviewed and approved the manuscript.

Declaration of interests

The authors declare no competing interests.

References

1. GBD 2019 Antimicrobial Resistance Collaborators. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* (2022) doi:10.1016/S0140-6736(22)02185-7.
2. von Gottberg, A. *et al*. Effects of vaccination on invasive pneumococcal disease in South

Africa. *N. Engl. J. Med.* **371**, 1889–1899 (2014).

3. Ganaie, F. *et al.* A New Pneumococcal Capsule Type, 10D, is the 100th Serotype and Has a Large *cps* Fragment from an Oral Streptococcus. *mBio* vol. 11 Preprint at https://doi.org/10.1128/mbio.00937-20 (2020).

4. Brueggemann, A. B. *et al.* Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J. Infect. Dis.* **187**, 1424–1432 (2003).

5. Grabenstein, J. D. & Musey, L. K. Differences in serious clinical outcomes of infection caused by specific pneumococcal serotypes among adults. *Vaccine* **32**, 2399–2405 (2014).

6. Müller, A. *et al.* Streptococcus pneumoniae Serotypes Associated with Death, South Africa, 2012-2018. *Emerg. Infect. Dis.* **28**, 166–179 (2022).

7. Cohen, C. *et al.* Streptococcus pneumoniae Serotypes and Mortality in Adults and Adolescents in South Africa: Analysis of National Surveillance Data, 2003 - 2008. *PLoS One* **10**, e0140185 (2015).

8. Donati, C. *et al.* Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. *Genome Biol.* **11**, R107 (2010).

9. Mostowy, R. J. *et al.* Pneumococcal Capsule Synthesis Locus cps as Evolutionary Hotspot with Potential to Generate Novel Serotypes by Recombination. *Mol Biol Evol* **34**, 2537–2554 (2017).

10. Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* **45**, 656–663 (2013).

11. Gladstone, R. A. *et al.* International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* **43**, 338–346

(2019).

12. Croucher, N. J. *et al.* Selective and genetic constraints on pneumococcal serotype switching. *PLoS Genet* **11**, e1005095 (2015).

13. Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* **6**, 288–301 (2008).

14. Hathaway, L. J., Grandgirard, D., Valente, L. G., Täuber, M. G. & Leib, S. L. Streptococcus pneumoniae capsule determines disease severity in experimental pneumococcal meningitis. *Open Biol.* **6**, (2016).

15. Hamaguchi, S., Zafar, M. A., Cammer, M. & Weiser, J. N. Capsule Prolongs Survival of Streptococcus pneumoniae during Starvation. *Infect. Immun.* **86**, (2018).

16. Im, H. *et al.* Anatomical Site-Specific Carbohydrate Availability Impacts Streptococcus pneumoniae Virulence and Fitness during Colonization and Disease. *Infect. Immun.* **90**, e0045121 (2022).

17. Arends, D. W. *et al.* Differential Pneumococcal Growth Features in Severe Invasive Disease Manifestations. *Microbiol Spectr* **10**, e0005022 (2022).

18. Hathaway, L. J. *et al.* Capsule type of Streptococcus pneumoniae determines growth phenotype. *PLoS Pathog.* **8**, e1002574 (2012).

19. Tóthpál, A., Desobry, K., Joshi, S. S., Wyllie, A. L. & Weinberger, D. M. Variation of growth characteristics of pneumococcus with environmental conditions. *BMC Microbiol.* **19**, 304 (2019).

20. Small, P. M., Täuber, M. G., Hackbarth, C. J. & Sande, M. A. Influence of body temperature on bacterial growth rates in experimental pneumococcal meningitis in rabbits.

Infect. Immun. **52**, 484–487 (1986).

21. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139**, 1366–1375 (2009).
22. Nilsson, G., Belasco, J. G., Cohen, S. N. & von Gabain, A. Growth-rate dependent regulation of mRNA stability in Escherichia coli. *Nature* **312**, 75–77 (1984).
23. Rogers, A. H., Pilowsky, K. & Zilm, P. S. The effect of growth rate on the adhesion of the oral bacteria Streptococcus mutans and Streptococcus milleri. *Arch. Oral Biol.* **29**, 147–150 (1984).
24. Russell, J. B., Sharp, W. M. & Baldwin, R. L. The effect of pH on maximum bacterial growth rate and its possible role as a determinant of bacterial competition in the rumen. *J. Anim. Sci.* **48**, 251–255 (1979).
25. Lin-Chao, S. & Bremer, H. Effect of the bacterial growth rate on replication control of plasmid pBR322 in Escherichia coli. *Mol. Gen. Genet.* **203**, 143–149 (1986).
26. Nabergoj, D., Modic, P. & Podgornik, A. Effect of bacterial growth rate on bacteriophage population growth rate. *Microbiologyopen* **7**, e00558 (2018).
27. Łapińska, U. *et al.* Fast bacterial growth reduces antibiotic accumulation and efficacy. *Elife* **11**, (2022).
28. Smirnova, G. V. & Oktyabrsky, O. N. Relationship between Escherichia coli growth rate and bacterial susceptibility to ciprofloxacin. *FEMS Microbiol. Lett.* **365**, (2018).
29. Tuomanen, E., Cozens, R., Tosch, W., Zak, O. & Tomasz, A. The rate of killing of Escherichia coli by beta-lactam antibiotics is strictly proportional to the rate of bacterial growth. *J. Gen. Microbiol.* **132**, 1297–1304 (1986).
30. Grant, A. J. *et al.* Bacterial growth rate and host factors as determinants of intracellular

bacterial distributions in systemic Salmonella enterica infections. *Infect. Immun.* **77**, 5608–5611 (2009).

31. Müller, A. *et al.* Pneumococcal serotype determines growth and capsule size in human cerebrospinal fluid. *BMC Microbiol.* **20**, 16 (2020).
32. Gulig, P. A. & Doyle, T. J. The Salmonella typhimurium virulence plasmid increases the growth rate of salmonellae in mice. *Infect. Immun.* **61**, 504–511 (1993).
33. Shlla, B. *et al.* The Rgg1518 transcriptional regulator is a necessary facet of sugar metabolism and virulence in Streptococcus pneumoniae. *Mol. Microbiol.* **116**, 996–1008 (2021).
34. Zhang, C. *et al.* MetR is a molecular adaptor for pneumococcal carriage in the healthy upper airway. *Mol. Microbiol.* **116**, 438–458 (2021).
35. Gao, Y. & Li, H. Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat. Methods* **15**, 1041–1044 (2018).
36. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, (2015).
37. Cremers, A. J. H. *et al*. Effects of 7-valent pneumococcal conjugate 1 vaccine on the severity of adult 2 bacteremic pneumococcal pneumonia. *Vaccine* **32**, 3989–3994 (2014).
38. Enright, M. C. & Spratt, B. G. A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. *Microbiology* **144 (Pt 11)**, 3049–3060 (1998).
39. Cremers, A. J. H. *et al*. The Contribution of Genetic Variation of Streptococcus pneumoniae to the Clinical Manifestation of Invasive Pneumococcal Disease. *Clin Infect Dis* **68**, 61–69 (2019).

40. Cremers, A. J. H. *et al*. The post-vaccine microevolution of invasive Streptococcus pneumoniae. *Sci Rep* **5**, 14952 (2015).

41. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).

42. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* vol. 401 877–884 Preprint at https://doi.org/10.1038/44766 (1999).

43. Jaillard, M. *et al.* A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet.* **14**, e1007758 (2018).

44. Lees, J. A. *et al.* Improved Prediction of Bacterial Genotype-Phenotype Associations Using Interpretable Pangenome-Spanning Regressions. *MBio* **11**, (2020).

45. Power, R. A., Parkhill, J. & de Oliveira, T. Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* **18**, 41–50 (2017).

46. Rodenburg, G. D. *et al.* Effects of pneumococcal conjugate vaccine 2 years after its introduction, the Netherlands. *Emerg. Infect. Dis.* **16**, 816–823 (2010).

47. Schaffner, T. O. *et al.* A point mutation in cpsE renders Streptococcus pneumoniae nonencapsulated and enhances its growth, adherence and competence. *BMC Microbiol.* **14**, 210 (2014).

48. Bentley, S. D. *et al.* Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**, e31 (2006).

49. Hanage, W. P. *et al.* Invasiveness of serotypes and clones of Streptococcus pneumoniae among children in Finland. *Infect. Immun.* **73**, 431–435 (2005).

50. Colijn, C., Corander, J. & Croucher, N. J. Designing ecologically optimized pneumococcal vaccines using population genomics. *Nat Microbiol* **5**, 473–485 (2020).

51. Massora, S. *et al.* Invasive disease potential of Streptococcus pneumoniae serotypes before and after 10-valent pneumococcal conjugate vaccine introduction in a rural area, southern Mozambique. *Vaccine* **37**, 7470–7477 (2019).

52. Rolfe, M. D. *et al.* Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol* **194**, 686–701 (2012).

53. Lees, J. A. *et al.* Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nat. Commun.* **10**, 2176 (2019).

54. Li, Y. *et al.* Genome-wide association analyses of invasive pneumococcal isolates identify a missense bacterial mutation associated with meningitis. *Nat. Commun.* **10**, 178 (2019).

55. Young, B. C. *et al.* Panton–Valentine leucocidin is the key determinant of Staphylococcus aureus pyomyositis in a bacterial GWAS. *eLife* vol. 8 Preprint at https://doi.org/10.7554/elife.42486 (2019).

56. Chaguza, C. *et al.* Prophage-encoded immune evasion factors are critical for host infection, switching, and adaptation. *Cell Genom* **2**, (2022).

57. Davies, M. R. *et al.* Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat. Genet.* **51**, 1035–1043 (2019).

58. Chewapreecha, C. *et al.* Genetic variation associated with infection and the environment in the accidental pathogen. *Commun Biol* **2**, 428 (2019).

59. Peters, S. *et al.* Campylobacter jejuni genotypes are associated with post-infection irritable bowel syndrome in humans. *Commun Biol* **4**, 1015 (2021).

60. Biggel, M. *et al.* Horizontally acquired papGII-containing pathogenicity islands underlie the emergence of invasive uropathogenic Escherichia coli lineages. *Nat. Commun.* **11**, 5968

(2020).

61. Wee, B. A. *et al.* Population analysis of Legionella pneumophila reveals a basis for resistance to complement-mediated killing. *Nat. Commun.* **12**, 7165 (2021).

62. Laabei, M. *et al.* Predicting the virulence of MRSA from its genome sequence. *Genome Res.* **24**, 839–849 (2014).

63. Chaguza, C. *et al.* Bacterial genome-wide association study of hyper-virulent pneumococcal serotype 1 identifies genetic variation associated with neurotropism. *Commun Biol* **3**, 559 (2020).

64. Boeck, L. *et al.* Mycobacterium abscessus pathogenesis identified by phenogenomic analyses. *Nat Microbiol* **7**, 1431–1441 (2022).

65. Galardini, M. *et al.* Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus Escherichia revealed by a genome-wide association study. *PLoS Genet.* **16**, e1009065 (2020).

66. Ma, K. C. *et al.* Adaptation to the cervical environment is associated with increased antibiotic susceptibility in Neisseria gonorrhoeae. *Nat. Commun.* **11**, 4126 (2020).

67. Clark, R. R., Lapierre, P., Lasek-Nesselquist, E., Gray, T. A. & Derbyshire, K. M. A Polymorphic Gene within the Mycobacterium smegmatis Locus Determines Mycobacterial Self-Identity and Conjugal Compatibility. *MBio* **13**, e0021322 (2022).

68. Chewapreecha, C. *et al.* Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *PLoS Genet.* **10**, e1004547 (2014).

69. Coll, F. *et al.* Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. *Nat. Genet.* **50**, 307–316 (2018).

70. Creasy-Marrazzo, A. *et al*. Genome-wide association studies reveal distinct genetic correlates and increased heritability of antimicrobial resistance in Vibrio cholerae under anaerobic conditions. *Microb. Genom.* **8**, (2022).

71. Farhat, M. R. *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nat. Genet.* **45**, 1183–1189 (2013).

72. Hicks, N. D., Carey, A. F., Yang, J., Zhao, Y. & Fortune, S. M. Bacterial Genome-Wide Association Identifies Novel Factors That Contribute to Ethionamide and Prothionamide Susceptibility in Mycobacterium tuberculosis. *MBio* **10**, (2019).

73. Farhat, M. R. *et al.* GWAS for quantitative resistance phenotypes in Mycobacterium tuberculosis reveals resistance genes and regulatory regions. *Nat. Commun.* **10**, 2128 (2019).

74. Alam, M. T. *et al.* Dissecting vancomycin-intermediate resistance in staphylococcus aureus using genome-wide association. *Genome Biol. Evol.* **6**, 1174–1185 (2014).

75. Mortimer, T. D., Zhang, J. J., Ma, K. C. & Grad, Y. H. Loci for prediction of penicillin and tetracycline susceptibility in Neisseria gonorrhoeae: a genome-wide association study. *Lancet Microbe* **3**, e376–e381 (2022).

76. Bertrand, R. L. Lag Phase Is a Dynamic, Organized, Adaptive, and Evolvable Period That Prepares Bacteria for Cell Division. *J Bacteriol* **201**, (2019).

77. Li, J. *et al.* Predictive genomic traits for bacterial growth in culture versus actual growth in soil. *ISME J* **13**, 2162–2172 (2019).

78. Höltje, J. V. & Tomasz, A. Lipoteichoic acid: a specific inhibitor of autolysin activity in Pneumococcus. *Proc Natl Acad Sci U S A* **72**, 1690–1694 (1975).

79. Mallawaarachchi, S. *et al.* Genome-wide association, prediction and heritability in bacteria

with application to. *NAR Genom Bioinform* **4**, lqac011 (2022).

80. Johnson, H. L. *et al.* Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Med.* **7**, (2010).

81. McGee, L. *et al.* Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. *J. Clin. Microbiol.* **39**, 2565–2571 (2001).

82. Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30**, 1224–1228 (2013).

83. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).

84. Hedge, J. & Wilson, D. J. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *MBio* **5**, e02158 (2014).

85. Tonner, P. D. *et al.* A Bayesian non-parametric mixed-effects model of microbial growth curves. *PLoS Comput. Biol.* **16**, e1008366 (2020).

86. Yang, E. *et al.* Influence of culture media, pH and temperature on growth and bacteriocin production of bacteriocinogenic lactic acid bacteria. *AMB Express* **8**, 10 (2018).

87. Alkema, W., Boekhorst, J., Wels, M. & van Hijum, S. A. F. T. Microbial bioinformatics for food safety and production. *Briefings in Bioinformatics* vol. 17 283–292 Preprint at https://doi.org/10.1093/bib/bbv034 (2016).

88. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

89. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16**, 1190–1208 (1995).

90. Argimón, S. *et al.* A global resource for genomic predictions of antimicrobial resistance and surveillance of Salmonella Typhi at pathogenwatch. *Nat. Commun.* **12**, 2879 (2021).

91. Epping, L. *et al.* SeroBA: rapid high-throughput serotyping of Streptococcus pneumoniae from whole genome sequence data. *Microb Genom* **4**, (2018).

92. Lees, J. A. *et al.* Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res.* **29**, 304–316 (2019).

93. Page, A. J. *et al.* : rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* **2**, e000056 (2016).

94. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

95. Lees, J. A. *et al.* Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation based study. *Wellcome Open Res* **3**, 33 (2018).

96. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

97. Argimón, S. *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* **2**, e000093 (2016).

98. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* vol. 3 217–223 Preprint at https://doi.org/10.1111/j.2041-210x.2011.00169.x (2012).

99. Holley, G. & Melsted, P. Bifrost: highly parallel construction and indexing of colored and

compacted de Bruijn graphs. *Genome Biol.* **21**, 249 (2020).

100. Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**, 4310–4312 (2018).

101. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Preprint at https://doi.org/10.1101/005165.

102. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

103. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

104. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

105. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

106. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models Usinglme4. *J. Stat. Softw.* **67**, (2015).

107. Croucher, N. J. *et al.* Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone Streptococcus pneumoniaeSpain23F ST81. *J. Bacteriol.* **191**, 1480–1489 (2009).

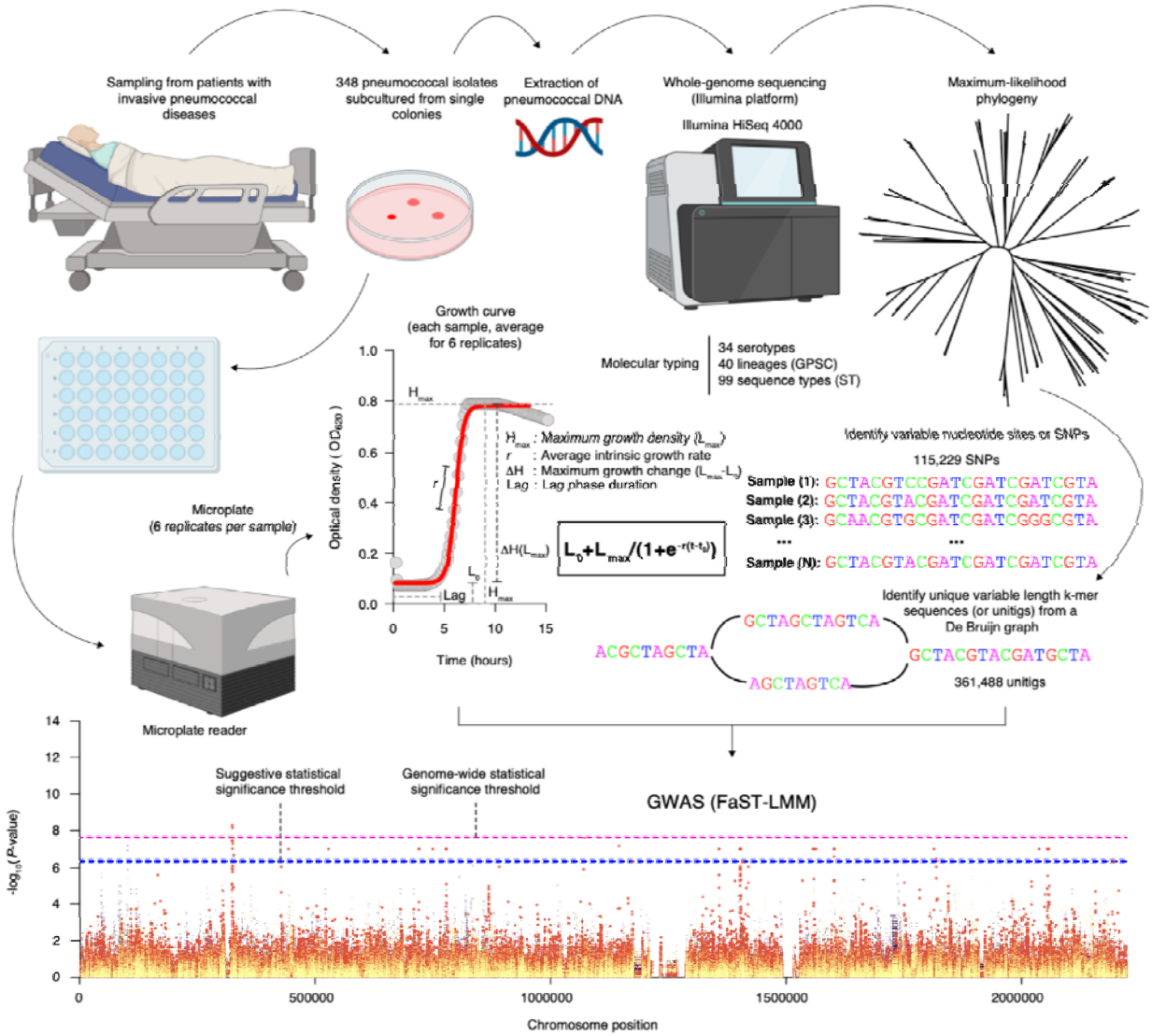

Fig. 1. Schematic overview of high-throughput quantification of *in vitro* pneumococcal growth kinetics and the genome-wide discovery of individual genomic variation influencing *in vitro* pneumococcal growth. Population genomic approaches using genome-wide association study (GWAS) offer a robust way to identify genomic variation influencing the growth dynamics of the pneumococcus. First, we collected 348 pneumococcal blood culture isolates that were obtained from 348 patients admitted to the hospital with invasive pneumococcal disease. The isolates were inoculated in a liquid medium in a highly standardised manner followed by measurement of the growth kinetics after every ten minutes for slightly over 15 hours using a microplate reader[17]. We then fitted a logistic function (see methods) to measure growth features, namely, the average growth rate (r), maximum growth density (H_{max}) defined as L_{max}, lag phase duration, and maximum growth change (ΔH), defined as the difference between the maximum

and minimum density (L_{max}-L_{min}). Additionally, genomic DNA was extracted from the cultured isolates for whole genome sequencing using the Illumina HiSeq 4000 sequencing platform. The generated sequencing data was used to reconstruct a maximum-likelihood phylogenetic tree of the isolates for downstream analyses, for example, the assessment of phylogenetic signals for the estimated growth features. We also performed *de novo* assembly of the sequencing reads to generate contigs, which were used to map against the ATCC 700669 pneumococcal reference genome (GenBank accession: NC_011900) to identify consensus single nucleotide polymorphisms (SNPs). We also generated presence, and absence patterns for unitigs or variable length *k*-mer sequences based on the draft assembled contigs. The SNPs were used to generate the phylogeny of the isolates while the unitigs were used in a GWAS to identify variants associated with the estimated pneumococcal growth features. The images were created with BioRender (https://www.biorender.com/).

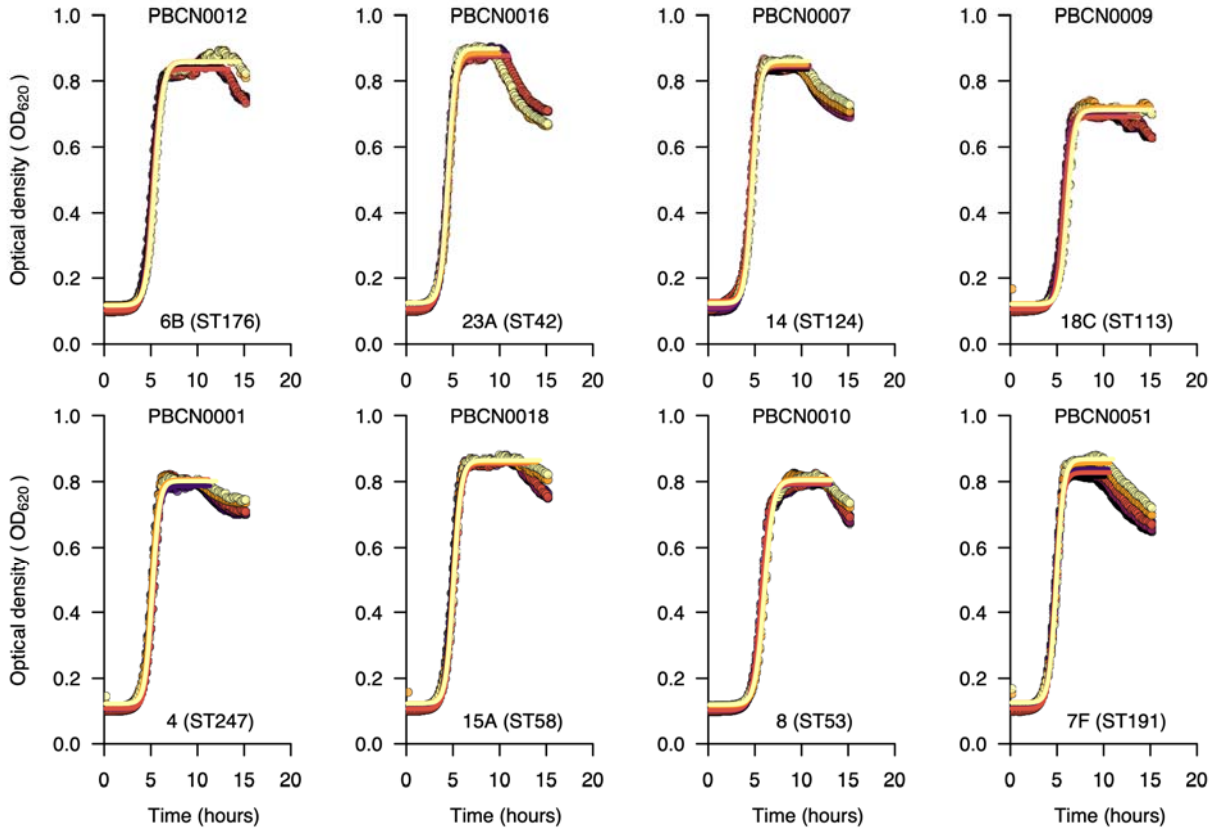

Fig. 2. Growth curves for selected serotypes show the *in vitro* pneumococcal growth dynamics and estimated growth features inferred by mathematical modelling. The growth curves above show the optical density at 620 nm monitored over 15 hours for an example dataset of 20 pneumococcal isolates representing different serotypes and sequence types (STs) based on the pneumococcal MLST scheme. We used the nonlinear optimisation algorithm to identify a combination of parameters with the best fit to a logistic curve based on the lowest residual sum of squares for the predicted value based on the model relative to the observed growth data (see methods). We fitted separate curves based on a subset of points for each replicate up to the end of the stationary phase or end of the experiment, whichever came first. We did this to get a better fit for the logistic curve during the growth phase, as it does not allow for negative growth, which happened after the stationary phase. We fitted the curves to infer the growth features for each of the six replicates. The averaged estimates for each replicate were then used for the downstream analyses. Additional plots showing fitted curves for all the isolates are shown in **Supplementary Fig. 1**.

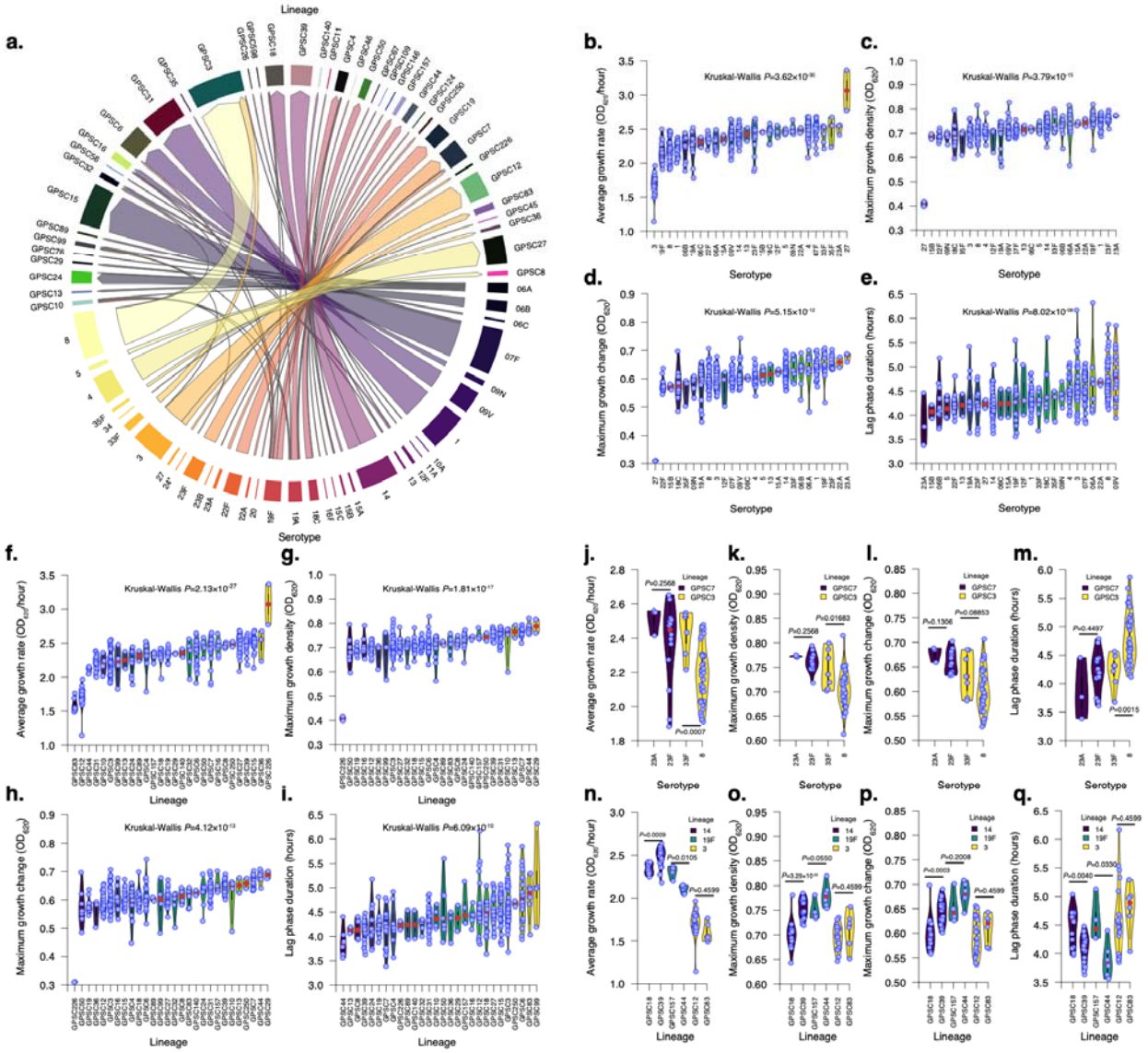

Fig. 3. The *in vitro* pneumococcal growth features show variable patterns dependent on the capsular serotype and genetic background or lineage. (a) Chord diagram showing the association between pneumococcal capsular serotypes and lineages for the Dutch isolates included in this study. The serotypes were determined using *in silico* approaches based on the sequencing data using SeroBA[91], while the lineages were defined using the PopPUNK framework[92] based on the global pneumococcal sequencing cluster (GPSC) lineage nomenclature[11]. The thickness of the connecting lines between serotypes and lineages in the chord diagram represents the percentage of isolates belonging to each serotype and lineage. Due to capsule switching, some serotypes were associated with multiple lineages, and similarly, some lineages were associated with multiple serotypes. The violin plot shows the distribution of the estimated average growth rate across different serotypes. The distribution of the estimated

growth features varied across serotypes with at least three isolates, as shown by the violin plots for the **(b)** average growth rate, **(c)** maximum growth change, **(d)** maximum growth change across serotypes, **(e)** lag phase duration growth features. Similarly, the distribution of the estimated growth features varied across lineages with at least three isolates, as shown by the violin plots for the **(f)** average growth rate, **(g)** maximum growth change, **(h)** maximum growth density, and **(i)** lag phase duration growth features. A comparison of the growth features between two serotypes belonging to the same lineages with at least three isolates is shown for the **(j)** growth rate, **(k)** maximum growth, **(l)** maximum growth change, and **(m)** the lag phase duration features. A comparison of the growth features between two lineages expressing the same serotype containing at least three isolates are shown for the **(n)** growth rates, **(o)** maximum growth, **(p)** maximum growth change, and **(q)** the lag phase features. Additional information regarding the isolates and their estimated growth features or parameters are available in **Supplementary Data 2**.

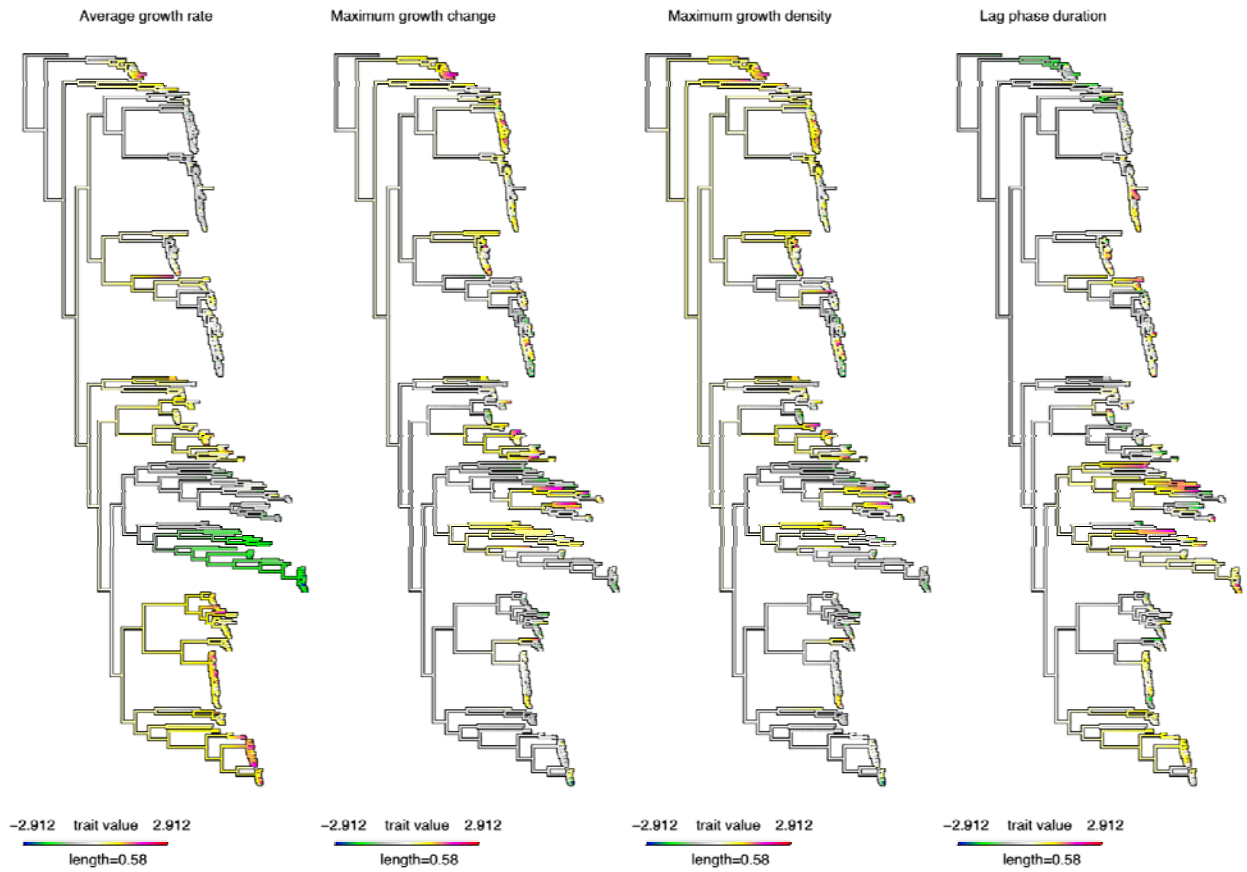

Fig. 4. Ancestral reconstruction reveals a strong correlation between the estimated *in vitro* pneumococcal growth features and the whole-genome phylogeny of the isolates. The evolution of the normalised estimated continuous growth features, namely average growth rate, maximum growth change, maximum growth density, and lag phase duration, were assessed by estimating the states of the internal nodes of the maximum likelihood phylogenetic tree of the isolates using a maximum likelihood ancestral reconstruction approach[98]. The length of the scale bar shows the number of nucleotide substitutions per site, while the colours represent the lowest and highest unnormalised growth features. The length of the scale bar shows the number of nucleotide substitutions per site, while the colours represent the lowest and highest unnormalised growth features. A phylogenetic tree showing bootstrap support values on the internal nodes of the tree is provided in **Supplementary Data 3**.

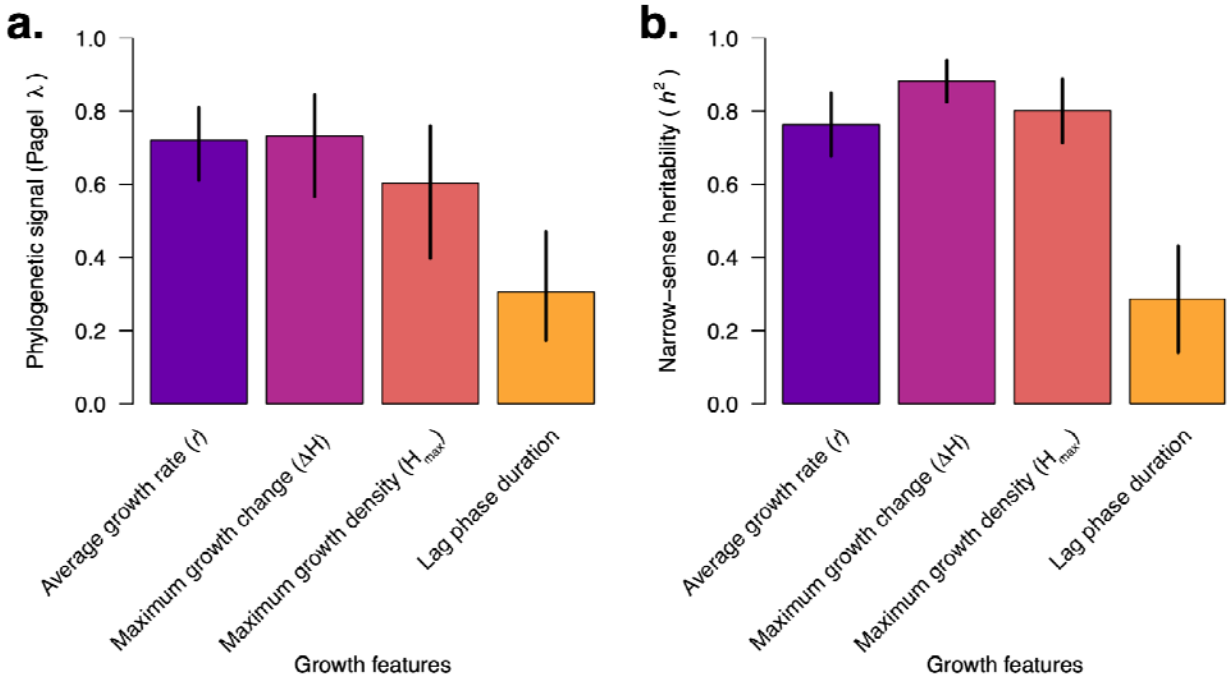

Fig. 5. The estimated *in vitro* pneumococcal growth features show high phylogenetic signals and narrow-sense heritability. (a) Bar plot showing the phylogenetic signal of the growth features using Pagel's λ statistic[42]. Pagel's λ measures the correlated evolution of the traits with the phylogenetic tree of the isolates. Pagel's λ estimates close to zero indicate no phylogenetic signal, i.e., the growth features were independent of the phylogeny, while values close to one show a strong phylogenetic signal. The average growth rate, maximum growth change, and maximum growth density showed the highest phylogenetic signal, while the lag phase duration feature showed an intermediate signal. In contrast, the maximum growth rate defined by Arends et al.[17] had the lowest phylogenetic signal, i.e., evolved more independently of the phylogeny. **(b)** Bar plot showing the narrow-sense heritability (h^2) of the pneumococcal growth features based on unitig sequences. The narrow-sense heritability corresponds to the amount of variability in the traits, i.e., growth features, explained by the pneumococcal genetic variation. The heritability is the proportion of phenotypic variation explained by the bacterial genetics (PVE) parameter using the linear mixed effects model implemented in GEMMA[102]. A heritability value close to zero implies low heritability or minimal contribution of genetics to the variability of the traits. In contrast, values close to one suggest a strong influence of genetics on growth features. There was a substantial impact of pneumococcal genetics on all the growth features, with the highest influence seen for the average growth rate, maximum growth change, and maximum growth density, with a slightly lower contribution for the lag phase duration and the maximum

growth rate measured by Arends et al.[17]. All the error bars represent the 95% confidence intervals. Additional information is provided in **Supplementary Data 3**.

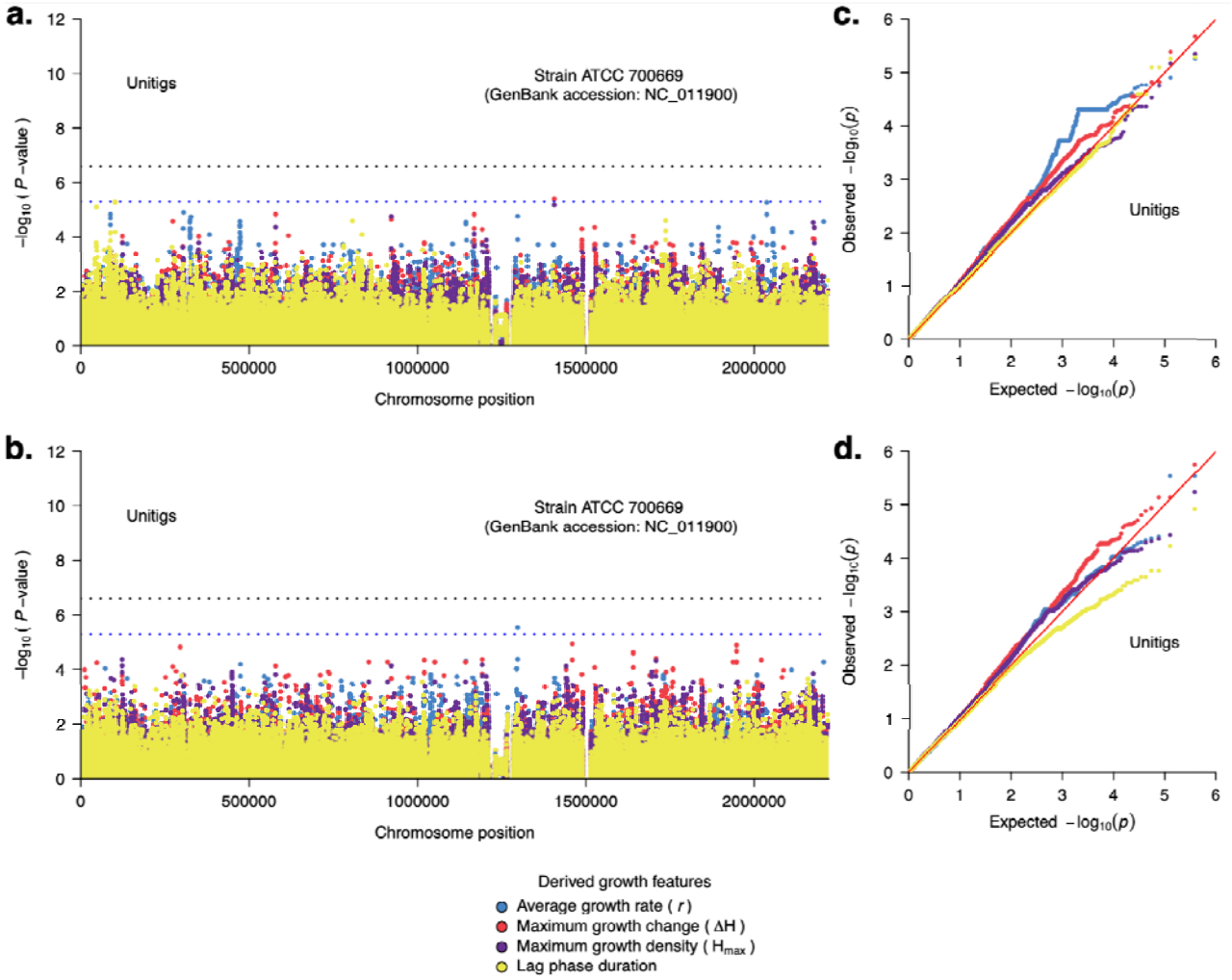

Fig. 6. Pneumococcal genetic variation associated with the *in vitro* growth features or parameters of the isolates. Manhattan plot showing the statistical significance of the hits from the GWAS analysis based on unitigs **(a)** without adjusting for the serotype and **(b)** after adjusting for the serotype. The statistical significance (*P*-value) were inferred based on the likelihood ratio test using Pyseer[100] mapped against the ATCC 700669 pneumococcal reference genome (GenBank accession: NC_011900[107]) to identify their location in the pneumococcal genome. The dots in the Manhattan plots with different colours represent GWAS for different growth features, as shown in the key. The red dotted line designates the genome-wide threshold for considering a variant as significantly associated, statistically, with the growth features. The blue dotted line represents the threshold for classifying variants as suggestively associated with the growth features. **(c)** Quantile–quantile plots showing the relationship between the observed statistical significance and the expected statistical significance for the GWAS of growth features not adjusting for the serotype. **(d)** Quantile–quantile (QQ) plots showing the relationship between

the observed statistical significance and the expected statistical significance for the GWAS of growth features after adjusting for the capsular serotype.

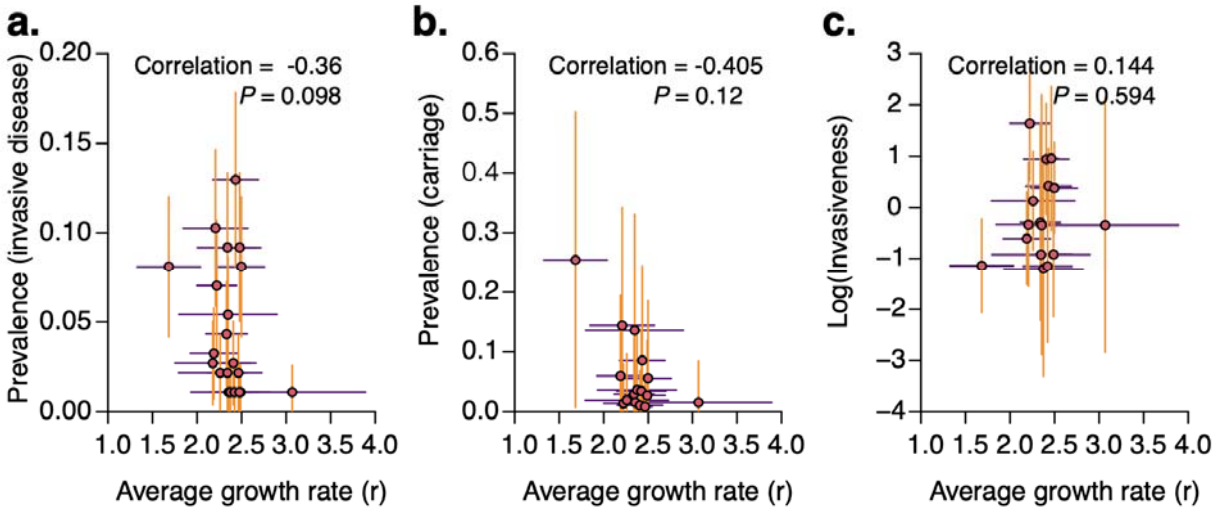

Fig. 7. Association between average in vitro pneumococcal growth rate with the prevalence of serotypes in invasive disease, nasopharyngeal carriage, and invasiveness. (a) Scatter plot showing the relationship between the prevalence of serotypes among invasive disease isolates collected before the introduction of PCV7 in the Netherlands and the average in vitro growth rate (r) of the isolates inferred as shown in Fig. 1. A Spearman correlation test showed no evidence of an association between the growth rate and prevalence of serotypes among the invasive disease isolates. **(b)** Scatter plot showing the relationship between the inferred prevalence of serotypes in the nasopharyngeal carriage before the introduction of PCV7 in the Netherlands and the inferred average in vitro growth rate. Due to the unavailability of the nasopharyngeal carriage data, we imputed the prevalence of the serotypes in carriage by dividing the prevalence in invasive disease by the average invasiveness of each serotype. The invasiveness index was calculated based on the meta-analysis of data collected before PCV7 introduction across multiple countries (see methods). A Spearman correlation test of the growth rate and prevalence of serotypes in carriage suggested a negative association between the variables. The association remained even after excluding serotype 3, which seemed like an outlier. **(c)** The relationship between the invasiveness and the average growth rates of each serotype from the pre-vaccination data. A Spearman correlation test of the growth rates and invasiveness suggested no association between the variables. All the error bars represent the 95% confidence intervals. Additional information is provided in Supplementary Data 3.